



# Tieteelliset lehdet ja takautuva digitointi

Digitointirahaa – onko sitä? -seminaari 11.6.2010  
Jyrki Ilva ([jyrki.ilva@helsinki.fi](mailto:jyrki.ilva@helsinki.fi))

# Digitointi ja tekstintunnistus

- Digitointia varten tarvitaan alkuperäiset lehdet
  - Onko omassa hyllyssä? Kirjaston kokoelmassa?
  - Pitääkö lehtien säilyä ehjänä?
- Suurempien aineistojen digitointi kannattaa ulkoistaa jollekin ammattimaiselle toimijalle
- Kannattaa tehdä kerralla hyvin, vaikka maksaisi vähän enemmän
  - Pitkäaikaissäilytys vs. käyttökopiot?
  - Kuvatiedostot vs. pdf-muoto
- Tekstintunnistuksen tekeminen (OCR) suositeltavaa
  - Tunnistettu teksti voidaan tallentaa pdf:n sisälle

# Kuvailutiedot

- Kuvaileva metadata
  - Tarvitaan esillepanoa varten
  - Julkaisun identifiointi ja tiedonhaun mahdollistaminen
  - Tekijä, nimeke, vuosikerta, numero, asiasanat, jne.
  - Kokoteksti-indeksointi voi korvata osittain sisällönkuvailua (mutta vain jos OCR tehty!)
- Tekninen ja hallinnollinen metadata
  - Tuotantoprosessin, käyttöoikeustietojen ym. dokumentointi mm. pitkäaikaissäilytyksen tarpeisiin
  - Voidaan tallentaa esim. METS-formaatissa
- Metadata saman tien myös kirjastojen järjestelmiin?

# Osaksi tiedonhaun infrastruktuureja

- Suositeltavaa tallentaa aineistot johonkin isompaan, vakiintuneeseen palveluun
  - Takaa pysyvän saatavuuden ja pysyvät osoitteet
  - Palvelun ylläpito ei kustantajan omalla vastuulla
  - Parantaa näkyvyyttä ja löydettävyyttä
  - Osaksi tiedonhaun infrastruktuureja, sinne missä tiedonhakijat muutenkin liikkuvat (esim. Arto, Nelli, KDK:n asiakasliittymä, Google, Google Scholar)
- Artikkelitason näkyvyys vs. numerot kokonaisuina?
  - Tiedonhakijat etsivät yleensä nimenomaan artikkeleita

# Bisnesmalli: maksullinen vai vapaa?

- Vapaa pääsy maksimoi käyttömahdollisuudet
  - Kuka tahansa mistä tahansa pääsee lukemaan
  - Etenkin kansainväliselle yleisölle suunnatulle, englanninkieliselle lehdelle vapaa pääsy järkevää
  - Toisaalta jonkun täytyy maksaa palvelun ylläpito
- Maksullisuudessakin on puolensa
  - Aineisto voi tuottaa säännöllisiä tuloja
- Oma palvelu tilaajille vai osaksi isompaa palvelua?
  - Käyttöoikeuksien valvonta aiheuttaa työtä
  - Kirjastot eivät tilaa yksittäisiä elektronisia lehtiä, ainoastaan isoja lehtipaketteja!

# Tekijänoikeuksien selvittäminen

- Jos julkaisuoikeuksista ei ole tekijän kanssa sovittu (yleensä ei ole!), oikeudet täytyy selvittää
  - Tekijänoikeus pääsääntöisesti voimassa 70 vuotta tekijän kuolemasta
  - Ei selkeää rajavuotta, suurin osa tieteellisten lehtien aineistosta tekijänoikeuden alaista
- Kopioston tarjoamat palvelut (Elektra ja Peri+)
  - Sidoksissa maksullisuuteen
  - Tulevaisuudessa sopimuslisenssit?
- Oikeudet voi selvittää myös itse
  - Työmäärä riippuu lehden koosta ja iästä

# Mitä Kansalliskirjasto voisi tarjota? (1)

- Digitointipalvelut (Mikkelin digitointikeskus)
  - Kehitetty digitointiprosessia, jossa tuotetaan samalla tarvittava hallinnollinen ja tekninen metadata
  - KK digitoi vuosina 2010-2011 kaikki suomalaiset aikakauslehdet vuoteen 1944 asti
  - Aineistojen saaminen laajempaan käyttöön?
- Metadataan liittyvät palvelut
  - Useimmat lehdet luetteloitu 80-luvulta asti Artoon
  - Ideoitu helppokäyttöistä syöttöjärjestelmää, jolla voisi tallentaa artikkelien kuvailutietoja myös mm. Artoon
  - Sovellettavissa myös takautuviin aineistoihin

# Mitä Kansalliskirjasto voisi tarjota? (2)

- Saatavuuteen ja löydettävyyteen liittyvät palvelut
  - Julkaisupalvelut (Elektra, Peri+)
  - Sekä maksullisuus että vapaa pääsy mahdollisia
  - Kytkenät tiedonhakuympäristöihin (Arto, Nelli, Kansallisen digitaalisen kirjaston asiakasliittymä)
  - Pysyvät tunnisteet ja osoitteet
  - Verkkoaineistojen pitkäaikaissäilytys joka tapauksessa KK:n lakisääteinen tehtävä
- Tekijänoikeuksiin liittyvät palvelut
  - Yhteistyö Kopioston kanssa (Elektra ja Peri+)